Brigham Young University
BYU ScholarsArchive

International Congress on Environmental Modelling and Software

7th International Congress on Environmental Modelling and Software - San Diego, California, USA - June 2014

Jun 16th, 10:40 AM - 12:00 PM

Towards an interoperable and distributed e-Infrastructure for Hydro-Meteorology: the DRIHM project

Antonella Galizia
IMATI-CNR, antonella@ge.imati.cnr.it

Daniele D'Agostino
IMATI-CNR, dago@ge.imati.cnr.it

Alfonso Quarti
IMATI-CNR, quarati@ge.imati.cnr.it

Gabriele Zereik
IMATI-CNR, zereik@ge.imati.cnr.it

Luca Roverelli
IMATI-CNR, roverelli@ge.imati.cnr.it

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

See next page for additional authors

Part of the Civil Engineering Commons, Data Storage Systems Commons, Environmental Engineering Commons, Hydraulic Engineering Commons, and the Other Civil and Environmental Engineering Commons

Galizia, Antonella; D'Agostino, Daniele; Quarti, Alfonso; Zereik, Gabriele; Roverelli, Luca; Danovaro, Emanuele; Clematis, Andrea; Delogu, Fabio; Parodi, Antonio; Harpham, Quillon; Jagers, Bert; Garrote, Luis; Dekic, Ljiljana; Dimitrijevic, Vladimir; Richard, Evelyne; and Caumont, Olivier, "Towards an interoperable and distributed e-Infrastructure for Hydro-Meteorology: the DRIHM project" (2014). *International Congress on Environmental Modelling and Software*. 6.
https://scholarsarchive.byu.edu/iemssconference/2014/Stream-B/6

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Presenter/Author Information

Antonella Galizia, Daniele D'Agostino, Alfonso Quarti, Gabriele Zereik, Luca Roverelli, Emanuele Danovaro, Andrea Clematis, Fabio Delogu, Antonio Parodi, Quillon Harpham, Bert Jagers, Luis Garrote, Ljiljana Dekic, Vladimir Dimitrijevic, Evelyne Richard, and Olivier Caumont

This event is available at BYU ScholarsArchive: https://scholarsarchive.byu.edu/iemssconference/2014/Stream-B/6

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Towards an interoperable and distributed e-Infrastructure for Hydro-Meteorology: the DRIHM project

Antonella Galizia, Daniele D'Agostino, Alfonso Quarati, Gabriele Zereik, Luca Roverelli, Emanuele Danovaro, Andrea Clematis, Elisabetta Fiori, Fabio Delogu, Antonio Parodi, Christian Straube, Nils Felde, Michael Schiffers, Dieter Kranzlmüller, Quillon Harpham, Bert Jagers, Luis Garrote, Vladimir Dimitrijevic, Ljiljana Dekic, Olivier Caumont, Evelyne Richard

IMATI-CNR, Genova, Italy {antonella,dago,quarati,zereik,roverelli,danovaro,clematis}@ge.imati.cnr.it

CIMA Foundation, Savona, Italy {elisabetta.fiori,fabio.delogu,antonio.parodi}@cimafoundation.org

LMU, Munich, Germany {straube,felde,schiffers}@nm.ifi.lmu.de, kranzlmueller@ifi.lmu.de

HR Wallingford, Wallingford, United Kingdom q.harpham@hrwallingford.com

DELTARES, Delft, Netherlands bert.jagers@deltares.nl

Technical University of Madrid, Spain l.garrote@upm.es

RHSS, {ljiljana.dekic,vladimir.dimitrijevic}@hidmet.gov.rs

CNRS, France Evelyne.Richard@aero.obs-mip.fr, olivier.caumont@meteo.fr

Abstract: Predicting weather and climate and its impacts on the environment, including hazards such as floods and landslides, is one of the main challenges of the 21st century with significant societal and economic implications. To advance the state of the art in forecasting extreme events, an effective collaboration with the Information and Communication Technology (ICT) scientific community is necessary to address challenges in combining meteorological, hydrological, hydraulic and impact observations and modelling tools seemingly and in a platform that allows scenario building and decision-support. With these issues in mind, the Distributed Research Infrastructure for Hydro-Meteorology (DRIHM) project aims at setting the stage for a new way of doing Hydro-Meteorological Research (HMR). The DRIHM goal is the development of an e-Science environment that allows users, from researchers to environmental agencies and also citizen scientists, to access and combine hydro-meteorological data and models using integrated specific services, user-friendly interfaces and proper computational infrastructures. This paper mainly focuses on the description of HMR requirements for the execution of complex HMR chains in an interoperable and distributed e-Infrastructure.

Keywords: e-Infrastructures; HM model simulations; interoperable HM workflows

1 INTRODUCTION

Forecasting severe storms and floods is a key topic in Environmental Modelling and could be considered, when related to extreme events, one of the main challenges of 21st century. Prediction of severe storms and (flash) floods is faced with hydro-meteorological and organizational challenges. The former is caused by the mandatory inclusion of several disciplines, e.g., atmospheric phenomena, cloud formations, flood impact assessment, which in turn requires the combined use of meteorological, hydrological, hydraulic, and impact modelling tools, a task which is far from reality. The latter addresses the urgent need of international collaboration in terms of data access, modeling, and expertise, since storms do not respect country boundaries.

The *Distributed Research Infrastructure for Hydro-Meteorology* (DRIHM) project[1] aims at tackling the aforementioned challenges by setting the stage for a new way of doing Hydro-Meteorological Research (HMR) Clematis et al. (2012). The project's main goal is the development of an e-Science environment facilitating the access and combination of hydro-meteorological data and models using integrated HMR services via user-friendly interfaces. In addition, the e-Science environment should be provided to a variety of user groups, ranging from HMR researchers to environmental agencies and even citizen scientists. In achieving this goal, DRIHM seamlessly accesses distributed computing, storage and networking facilities, as those provided by available European e-Infrastructure ecosystems such as *European Grid Infrastructure* (EGI)[2] and the *Partnership for Advanced Computing in Europe* (PRACE)[3].

However, there are two main roadblocks on the way towards the introduced DRIHM goal, i.e., *model coupling* and *high computing demands*.

Facilitating the above motivated combination of tools and models from multiple disciplines results in simulations consisting of and requiring the execution of several models organized in a workflow. A workflow, in turn, is implemented by **model coupling**, i.e., using the output of a model as input for the successive one. This can lead to compatibility issues among hydro-meteorological models at different levels. The models concatenated in a workflow in fact may depend on different execution environments, on organizational constraints, and on incompatible data formats and semantics to be bridged. Up to now, HM researcher used to hard-wire a reduced set of hydro-meteorological models, mainly due to not standardized interfaces and to difficult access to computational resources.

In addition, already a single hydro meteorological simulation or model of a workflow is rather complex and pose **high computing demands**. Furthermore, their execution is extremely time consuming and requires a strong, reliable, and capable IT background. Finally, HMR models usually have significant HPC requirements that pose specific constraints on the eligible resources needed for their efficient execution. Moreover, in general, a strong connection and collaboration with the *Information and Communication Technology* (ICT) scientific community could be very fruitful to provide new technological solutions, as recognized by Shapiro et al. Shapiro et al. (2007) and Shukla et al. Shukla et al. (2009, 2010).

This paper presents the DRIHM approach how to address the itemized roadblocks: model coupling, platform/computational requirements and the DRIHM portal. The first concerns the effort spent in model standardization in terms of data formats and semantics thus to easily compose models within a workflow. The second regards the design and the development of the DRIHM Distributed Computing Infrastructure (DDCI), i.e. the e-Infrastructure to effectively respond to the heterogeneous requirements posed by the models provided in the project. Finally, the solutions designed and implemented are provided to the users through the DRIHM portal, i.e. the user-friendly interface developed to access the services and resources of the HMR community. The portal shields users from the underlying infrastructure complexities and specific implementations, and lets the user to compose models in a desired workflow. The paper is organized as follows: in the next Section we discuss the issues related with model coupling and proposed solutions. Section 3 introduces the design of the DDCI, while Section 4 briefly discusses the DRIHM science gateway. Section 5 presents concluding remarks.

[1]The DRIHM project, 2011-2015, funded by the EU FP7, http://www.drihm.eu
[2]http://www.egi.eu
[3]http://www.prace-ri.eu

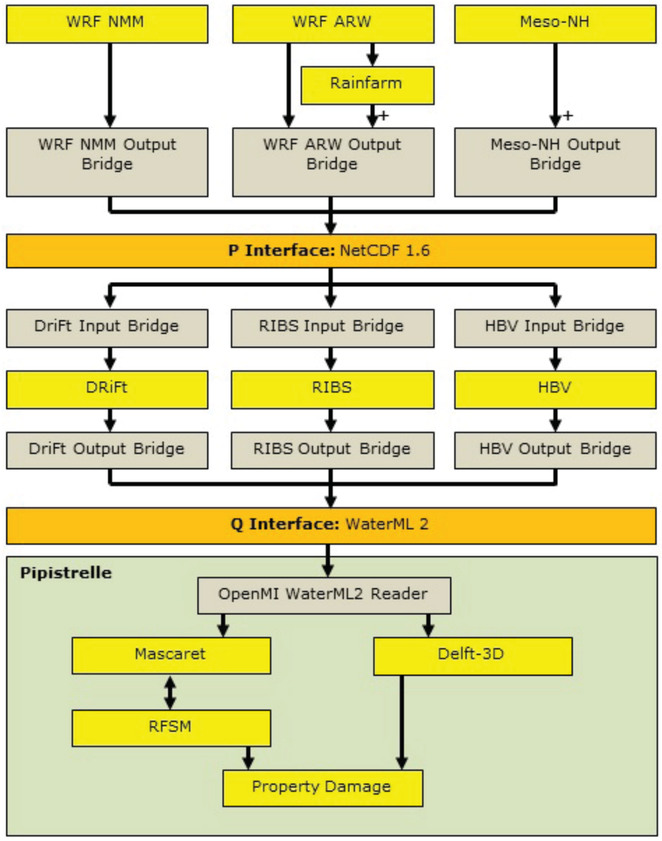


Figure 1. Models provided by the DRIHM e-Science environment

2 Enabling HMR Forecasting Chains through Standardized Interfaces

DRIHM aims at creating an easy-to-use e-Science environment that enables model execution *as a service* as well as the creation of (arbitrary) model chains. Both functionalities contribute to DRIHM's final objective, i.e. enabling the modeling, definition, and execution of a probabilistic forecasting chain on an existing e-Infrastructure.

A forecasting chain covers a variety of physical phenomena, ranging from atmospheric cloud formations to the assessment of societal impact created by the flood events. Hence, a full forecasting simulation run requires the execution of a (complex) workflow that is built of several models from different scientific domains. Enabling (model) concatenation imperatively requires addressing interoperability issues posed by HMR models Viglione et al. (2010), since each model needs to interact and exchange data and parameters with connected models within the workflow. Many attempts have been made in different countries to build up a sound HM probabilistic chain, notably in Southern Europe and in US, as described by Apel et al. (2006); Pappenberger et al. (2011). However, HM researchers still used to hard-wire a reduced set of Hydro-Meteorological models. In contrast, DRIHM aims at developing standardized interfaces to connect related models based on existing standards. A DRIHM forecasting chain corresponds to four model domains exposing differing technical and functional characteristics:

- Meteorological models simulating a variety of atmospheric phenomena including rainfall; the output data are 2D or 3D fields (rainfall, temperature, pressure, ...);
- Hydrological models simulating catchment runoff; input is rainfall (as 2D field and/or as time series - from rain gauge stations or ground radars); the output is a set of time-series describing streamflow at

the outlet of several sub-basins;

- Hydraulic models simulating movement of the water across land surfaces and along open channels; the input are time series describing surface overland flow entering the channels;
- Flow and Impact models producing flood damage results.

Summarized, this situation requires models being interoperable and the final solution being extensible and providing high usability. Executing model chains, and the ability to compare different models at each level, will require:

- standardization of interfaces (file format) among models composing the chain: the Precipitation ("P") Interface, between meteorological and hydrological models, the Discharge ("Q") Interface between hydrological drainage models and hydraulic open channel models, Hydraulic Interfaces between hydraulic models;
- standardization of meta-tags associated with input and output, to allow checking the chain composability before running the model chain.
- consistency check (i.e. meteorological simulation is performed on a domain including the drainage basin adopted in the following step).

Figure 1 presents the models that will be provided by the DRIHM e-Science environment, the developed "bridge" components and the designed interfaces. For each layer of the DRIHM chain, the state of the art modelling techniques will be utilized; in Figure 1 they are depicted with yellow blocks. We design standardized interfaces to bridge the models belonging to the different layers composing a workflow, and a set of conditions has been established at each interface; they are depicted with orange blocks in Figure 1. In fact, to be composed in a workflow in the DRIHM e-Science environment, HMR models have to be able to ingest input data and to produce output data complaint to the adopted standard. To this end, in some cases, the so-called "bridge" components have been developed; they are depicted with grey blocks in Figure 1. These components have to deal with specific standards like NetCDF-CF, used to encode gridded data type, and WaterML 2.0, used to encode time series. In the case of Hydraulic and Impact models, usually tightly coupled, we exploit standard available solutions already wide-spread in the scientific community as Open Model Interface (OpenMI)[4].

More specifically, a component named *Meteorological Model Bridge* (MMB) aimed at allowing ingestion and interoperability between meteorological models, such as WRF, Rainfarm,and Meso-NH, and hydrological models, such as DRiFt or RIBS, has been developed. Starting from the gridded outputs produced by meteorological models MMB extracts only the relevant output fields (mainly temperature and precipitation), and re-encoding the output according to the netCDF-CF convention. Moreover, a component named *Data Observations Ingestion* (DOI) has been designed for downloading observed point data as WaterML 2.0 data from the services available from the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI)[5] in San Diego (or any other instance of HydroServer). Such component provides observed data to hydro models (e.g. DRiFt, RIBS).

Beside interface format standardisation, we have to check that numerical simulations composed in a chain are compatible: the geographical domain of output data has to contain the domain specified in the required input data of the following model.

In order to check that the chain composition is meaningful, before submitting the jobs on the underlying infrastructure, we associate to each model a set of metadata, and check consistency against such metadata. We started from ISO19115 metadata standard for describing spatial datasets, and adopted FluidEarth extension to ISO19115. This allow to provide general information (such as a title, an abstract, owning organisation or contact details, execution platform, link to sources or binaries) but also, for each input and output: Name, Description, Format, whether it is mandatory. In Harpham et al. (2014) we further extended the set of metadata, including:

[4] http://www.openmi.org
[5] http://www.cuahsi.org

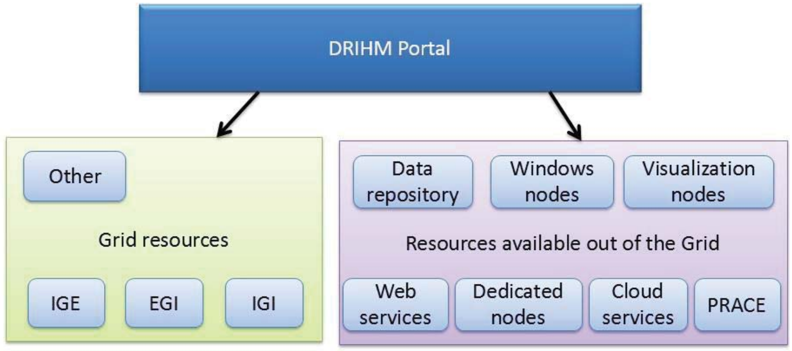


Figure 2. Schematic vision of the DRIHM Distributed Computing Infrastructure

- File - full pathname of the input/output file (if any);
- Feature Type - a description of the spatial/temporal structure of the data, such as PointSeries, GridSeries, . . .;
- Position - the 2-dimensional geospatial position of the data given as a rectangular bounding polygon;
- Time Range - the timestamp of the first (earliest) and last (latest) reading in the timeseries;
- Timestep Type - indicator of regular or irregular timestep interval;
- Maximum Timestep Interval - the length of the largest timestep represented in the data and its unit of measurement;
- Minimum Timestep Interval - the length of the smallest timestep represented in the data and its unit of measurement;
- Parameter Name and Unit - the name and unit of measurement of the physical parameter/phenomena represented.

3 The Design of a Distributed Computing Infrastructure for HMR

At the base of the DRIHM project there is the need to access distributed computing, storage and networking facilities thus to provide end user with a seamless environment able to carry on her/his advanced research experiments and simulations. DRIHM is not based on a proprietary or pre-defined computing infrastructure, but it aims to exploit the available European e-Infrastructure ecosystems. The architecture of the DRIHM Distributed Computing Infrastructure - DDCI is presented in Figure 2. Actually, models provided in the DRIHM e-Science environment differ in terms of platform requirements: some models are designed for POSIX systems, others are designed for Windows, some depend on specific libraries (sometimes even on specific library versions). Moreover the model coupling activities clearly highlighted specific requirements to fully and efficiently integrate the models in the computing infrastructure at the base of DRIHM.

First, it is necessary to distinguish between resources shared and accessible through the Grid, and resources available out of the Grid. As regards the Grid infrastructures, the DRIHM resource pool is provided by National Grid Initiatives through the EGI, in particular a strict interaction is undergoing with the Italian Grid Infrastructure. The DRIHM project considers also the resources from the Initiative for Globus in Europe (IGE)[6]; it has been used as a testbed to deploy preliminary versions of software tools and to test the interaction with infrastructure services, before the production deployment. Beside the Grid, there is the need to consider several "components" granting services that did not find place (for performance and/or functionality reasons) in the core Grid resources. In particular DRIHM considers the following resources:

[6]http://www.ige-project.eu recently became the European Globus Community Forum, http://www.egcf.eu

- PRACE resources to fulfill the computational requirements of meteorological models;
- dedicated nodes to run/provide models with specific constraints, e.g. compilers, libraries, Operative Systems, large databases;
- Windows resources to run the Windows-based models (e.g. Impact models);
- data repository to store/provide large data sets, as global circulation models;
- specialized capabilities for specific services, as remote visualization;
- Cloud services to provide service/resources at different levels;
- Web services to access and provide models and facilities out of the Cloud and Grid.

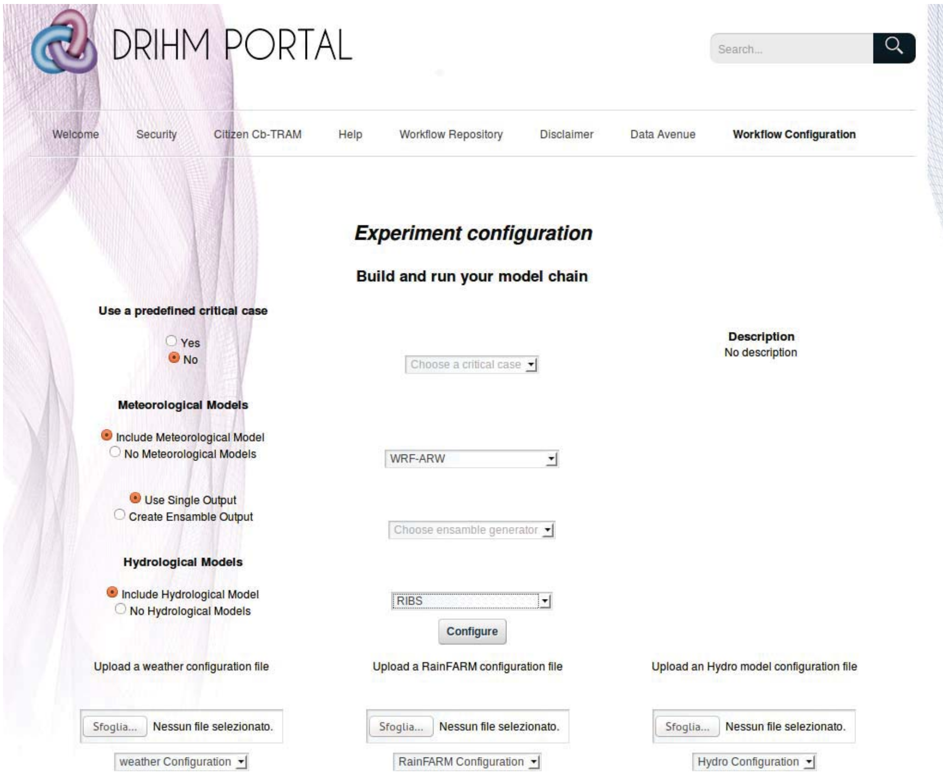
Furthermore, it is of paramount interest to adhere to the principle of transparency of the underlying distributed computing systems. Consequently, HMR models, tools and data sets need to be deployed in such a way that both location and migration transparency is achievable for data, models and workflows. Because of the different requirements, adaptations are inevitable (a task called "gridification"). Once all components have been "gridified", HMR applications may be composed and executed on distributed resources. A central DRIHM binary repository holds the gridified Linux-based model binaries, a second repository holds static data for specific models.

4 The DRIHM science gateway for HMR

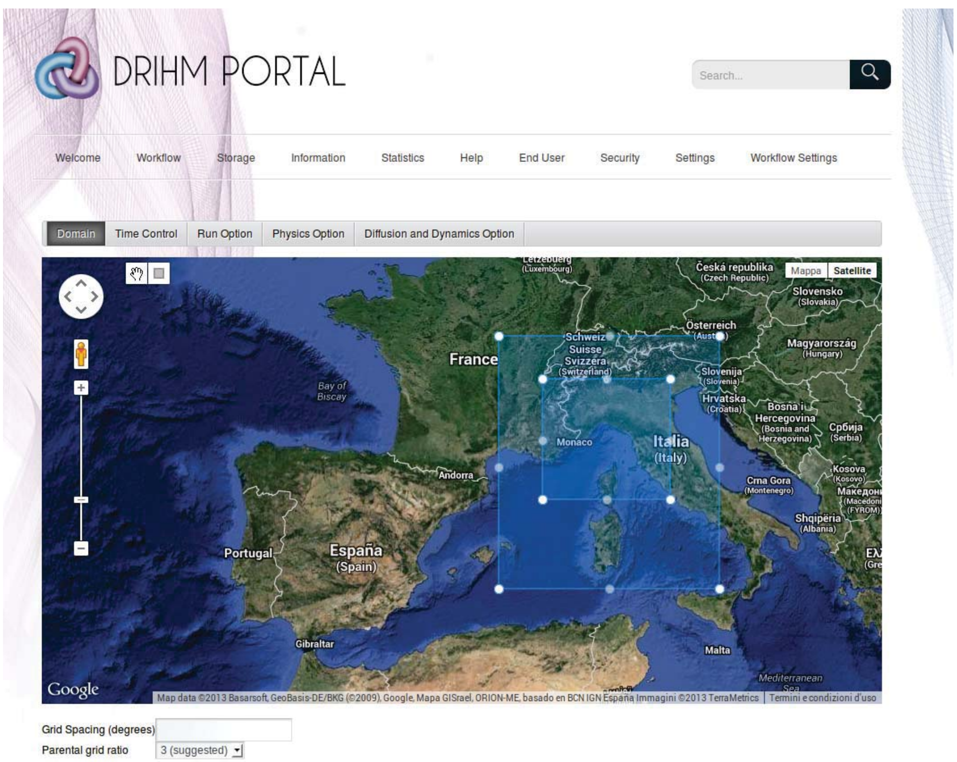
The DRIHM Portal is the scientific gateway designed to shape the DRIHM vision. The portal provides integrated solutions to manage and exploit the DDCI at different levels: users, resources, workflows, data, and services; it is more than a simple web user interface. The portal is based on the technologies proposed by the *SCIentific gateway Based User Support* (SCI-BUS) project[7], i.e. a customized version for e-Science environments of the generic-purpose gUSE/WS-PGRADE portal family. The portal, available at http://portal.drihm.eu, is currently in a development phase and new features are incrementally added through the development of specific portlets. The DRIHM Portal supports HM researchers in the design of new experiments. It exposes simple web-based interfaces that let the user to compose the desired workflow, to specify model parameters and finally to submit jobs. This corresponds to the development of a set of customized portlets to enable the workflow configuration. The portal generates all configuration files and handles the execution on the best suited computational resources. It manages the storing, seeking and downloading of the configurations files previously created.

Figure 3(a) shows an example of the workflow configuration portlet that enables the definition of a chain for HM simulations. A workflow can contain one or more models among those available; for each model composing the chain, a web interface (see Figure 3(b)) lets the user select the proper configuration parameters (e.g. river basin, microphysics, time and spatial domains). The ability to combine different models in a chain and to easily swap among analogous models required the large effort in model standardization previously described. Moreover it is worth to underline that the execution of the MMB and DOI bridge components, when necessary, is automatically performed as part of the workflow. The DRIHM portal takes care of the exploitation of the DDCI hiding all complexity to the final user. The interoperability among different e-Infrastructures exploits the gUSE/WS-PGRADE technologies enabling the integration of resources and services from Grid and Cloud infrastructures.

[7]https://www.sci-bus.eu



(a)



(b)

Figure 3. (a) DRIHM portal interface to workflow configuration (b) DRIHM portal interface to select the river basin

5 Conclusions

The DRIHM project aims at giving answers to the HMR issues related to the execution of complex and demanding simulations through the set up of an interoperable and distributed e-Infrastructure. The DRIHM e-Science environment is essentially a collection of components, some ICT and some HMR. Therefore the design is driven by the capabilities of these components; consequently the majority of the deployed services are not designed from scratch, but they are obtained through the integration of a set of already existent components (i.e. HMR models, middleware frameworks, third parties web applications). In this paper, we presented three specific aspects addressed by DRIHM: the horizontally model coupling to arrange the concatenation of models in a forecasting chain; the design of a distributed infrastructure to effectively respond to the heterogeneous requirements posed by the models provided in the project; the DRIHM portal to enable the user to access and exploit the e-Infrastructure. Initial experimentations to simulate specific critical cases defined in the project already gave scientific valuable results as presented by Caumont et al. (2014). The portal is further evolving with a new set of user-driven requests: in-place analysis of simulation results and ability to share experiments within the HMR community. Furthermore, in a companion EU FP7 research project DRIHM2US[8] we are investigating interoperability issues related to the use of XSEDE, eXtreme Science and Engineering Discovery Environment[9].

References

Apel, H. Thieken, A. Merz, B. and Blosch, G. (2006). A probabilistic modelling system for assessing flood risks. *Natural Hazards*, 38(12):79–100.

Caumont, O. and the DRIHM Experiment Suites Team (2014) Multi-model ensemble hydrometeorological modelling of the 4 November 2011 Genoa, Italy flash flood in the framework of the DRIHM project In *Geophysical Research Abstracts, EGU2014-11409*.

Clematis, A. D'Agostino, D. Danovaro, et al. (2012). DRIHM: Distributed Research Infrastructure for Hydro-Meteorology. In *Proceedings of the 7th International Conference on System of Systems Engineering SOSE 2012*, pages 149–154.

Pappenberger, F. Thielen, J. and del Medico, M. (2011). The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System. *Hydrological Processes*, 25(7):1091–1113.

Shapiro, M. A. Shukla, J. Hoskins, et al. (2007). The socio-economic and environmental benefits of a revolution in weather, climate and Earth-system analysis and prediction. *The Full Picture*, Group on Earth Observations, Tudor Rose, pages 137–139.

Shukla, J. Hagedorn, R. Hoskins, B. Kinter, J. Marotzke, J. Miller, M. Palmer, T. N. and Slingo, J. (2009). Revolution in climate prediction is both necessary and possible: A declaration at the World Modelling Summit for Climate Prediction. *Bull. Amer. Meteor. Soc.*, 90:175–178.

Shukla, J., Palmer, T. N. Hagedorn, R. Hoskins, B. Kinter, J. Marotzke, J. Miller, M. and Slingo, J. (2010). Toward a new generation of world climate research and computing facilities. *Bull. Amer. Meteor. Soc.*, 91:1407–1412.

Viglione, A. Borga, M. Balanbanis, P. and Bloschl, G. (2010). Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy. *J. Hydrol.*, 394:63–77.

Harpham, Q. Danovaro, E. (2014). Standard metadata to support models and interfaces in an hydrometeorological model chain. To be submitted *J. Hydrol.*.

[8] DRIHM2US project, 2012-2014, funded by the EU FP7, http://www.drihm2us.eu
[9] https://www.xsede.org